

A.Yousfi⁽¹⁾, N. Tazzite⁽²⁾, El.Bouyakhf⁽²⁾

⁽¹⁾Institut d'Etudes et de Recherches pour l'Arabisation, Université Mohamed V Souissi, Rabat, Maroc

⁽²⁾Faculté des Sciences, Université Mohamed V Agdal, Rabat, Maroc

yousfi71@yahoo.com,
naimaitm2004@yahoo.fr

Conception et réalisation d'un moteur de recherche arabe sur Internet

1- Introduction

Les moteurs de recherche sur Internet sont des programmes accessibles aux utilisateurs du réseau qui consultent une gigantesque base de données documentaire, où est enregistrée pour chaque mot-clé présent dans les pages d'Internet, une liste d'adresses de pages contenant ce mot-clé [ref0]. Un groupe de telles listes est nommé « index ». La création et la maintenance de ces index sont des tâches très lourdes, et c'est toujours un problème redoutable de déterminer quelle est l'information pertinente à retourner en réponse à la requête d'un utilisateur.

Les moteurs de recherche sur Internet en arabe sont de plus en plus nombreux, Nous citons par exemple :

- **Arabvista** : C'est l'un des moteurs de recherche les plus importants dont la spécificité est la possibilité offerte au client d'affiner sa recherche à partir des mots clé, critère de sa recherche. Il a été réalisé par la société Emirates Internet et Multimédia (EIM) et Compaq sur le modèle du site Altavista de Compaq. Il possède un dictionnaire arabe anglais et permet ainsi à ses utilisateurs de rechercher des informations en anglais sur le net à partir de l'arabe [ref1].
- **Tarjim** : C'est l'un des moteurs de recherche les plus puissants disponible sur le net. C'est un directory qui présente une liste de répertoire desquels le client effectue sa recherche [ref2].
- **Ayna** : Ayna est actuellement l'un des meilleurs moteurs de recherche en langue arabe sur Internet. Il est au service de plus de 280 millions de clients arabophones potentiels du Moyen-Orient, du Maghreb ainsi des internautes arabophones expatriés. Ayna est pour le monde arabe ce qu'est Yahoo pour l'occident [ref3].
- **Arabia** : Arabia est un portail qui offre plusieurs services [ref4]:
 - forums de discussion;
 - services d'information;
 - moteur de recherche

Un moteur de recherche se compose en général des modules suivants : un Robot, un programme indexeur, une base de données, un processus d'interrogation, une interface entre le moteur de recherche et l'utilisateur.

Dans ce travail, nous avons élaboré un moteur de recherche pour la langue arabe qui permet de faire une recherche à partir d'une requête sur des pages Web arabes. Ce moteur est constitué de deux modules :

- Module d'indexation : il permet d'analyser chaque document de la collection et d'extraire un ensemble de mots clés (termes) associé à chaque document. Chaque terme est associé à un ensemble d'informations qui sont stockées dans un index pour faciliter la recherche. L'extraction des termes est faite en se basant sur la fréquence d'apparition des mots dans chaque document.
- Module de recherche : ce module reçoit une requête (q). La recherche est faite en se basant sur la mesure de similarité entre cette requête et chaque document (d). Dans notre cas nous avons utilisé la mesure suivante :

$$\text{Sim} (d, q) = \sum_i (p_i * q_i)$$

Ensuite les résultats de recherche (documents) sont triés en ordre décroissant selon le degré de similarité entre les documents trouvés et la requête (q).



Interface pour la recherche

Conclusion et perspectives

Dans la suite de notre travail, on va intégrer l'analyse morphologique dans l'étape d'indexation et de la recherche pour réduire la taille de l'index d'une part et d'autre part de retourner les documents contenant aussi les dérivés des mots à rechercher.

Bibliographie

[ref0]: http://www.journaldunet.com/encyclopedia/definition/97/45/21/moteur_de_recherche.shtml

[ref1]: <http://www.arabvista.com/>

[ref2]: <http://www.tarjim.com/>

[ref3]: <http://www.ayna.com/>

[ref4]: <http://www.arabia.com/arabic>

[V.Combet]:Indexation automatique:un état de l'art.

[Mabrouka et Hachani]:Indexation automatique.

[Jacques Savoy] :Recherche d'information dans des corpus plurilingues.

[Claude de Loupy]: Évaluation de l'Apport de Connaissances Linguistiques en Désambiguïsation Sémantique et Recherche Documentaire